

Single and multiple error state-space models for signal extraction

José Casals^a

Sonia Sotoca^a

Miguel Jerez^{a*}

^a Departamento de Fundamentos del Análisis Económico II. Facultad de Ciencias Económicas.
Campus de Somosaguas. 28223 Madrid (SPAIN)

Abstract: We compare the results obtained by applying the same signal extraction procedures to two observationally equivalent state-space forms. The first model has different errors affecting the states and the observations, while the second has a single perturbation term which coincides with the one-step-ahead forecast error. The signals extracted from both forms are very similar but their variances are drastically different, because the states for the single-source error representation collapse to exact values while those coming from the multiple-error model remain uncertain. The implications of this result are discussed both, with theoretical arguments and practical examples. We find that single error representations have advantages to compute the likelihood or to adjust for seasonality, while multiple error models are better suited to extract a trend indicator. Building on this analysis, it is natural to adopt a ‘best of both worlds’ approach, which applies each representation to the task in which it has comparative advantage.

Keywords: State-space model; Signal extraction; Seasonal adjustment; Kalman filter; Trend

MSC classification: 60G35, 93E11, 93E14, 91B84

* Corresponding author. Email: mjerez@ccee.ucm.es, tel: (+34) 91 394 23 61

1. Introduction

In time series analysis, signal extraction is an important sub-field dealing with the definition and estimation of interesting signals buried in a given time series, such as its trend, cycle or seasonal component. State-Space (SS) methods are a natural choice for this task because of two reasons. First, the SS model represents the observed time series as a linear combination of latent variables, known as ‘states’, which may have very flexible dynamic and stochastic properties and are, therefore, naturally close to the idea of ‘buried signals’. Second, after collecting a sample and defining a suitable model, the SS literature provides robust and efficient algorithms to obtain estimates of the states, as well as the corresponding variances and covariances. Two popular choices to perform this numerical signal extraction are the Kalman filter (hereafter, KF) and the fixed-interval smoother [1].

This paper builds on the well-known fact that there are many observationally equivalent SS forms realizing the same output, each of them with its own pros and cons for specific uses. This idea has been used in many works. For example, [13] highlights the advantages of minimal representations, [19] shows the benefits of using canonical forms, or [21] applies it to the estimation of a class of nonlinear models.

On this basis, we compare the results obtained by applying the same state-estimation algorithms to two different linear and time-invariant SS forms. The first model has different errors affecting the states and the observations, while the second has a single perturbation term which coincides with the one-step-ahead forecast error. We will refer to these models as Multiple Error Model (MEM) and Single Error Model (SEM), respectively. In [6] we showed that any MEM can be written in an observationally equivalent SEM form, so both representations are equally general.

The advantages of the SEM form have been discussed in the literature referring to particular formulations.

De Jong and Penzer [12] point out that SEM models have been overlooked in time series literature and summarize its advantages for ARIMA modeling in the following terms: ‘The form has both, computational and conceptual advantages. With this form, the KF collapses, after the

processing of an initial stretch of data, to computing the exact moving average errors... [and] clarifies smoothing, maximum likelihood estimation and transformation issues.’

Other works concentrate in the computation of the Beveridge-Nelson decomposition with MEM and SEM forms ([2], [20] and [21]) and discuss their pros and cons with divergent results: while [2] and [21] highlight the advantages of the SEM form, [20] provides empirical evidence in favor of the MEM alternative. In fact, in this work we support the idea that the MEM and SEM alternatives have specific advantages for different applications.

As expected, signals extracted from equivalent MEM and SEM forms are very similar. However, their variances are drastically different because the states for the SEM collapse to values with zero variance. In previous works, we exploited this important property to improve the computation of the Gaussian likelihood, see [6], and to perform the trend-cycle-seasonal decomposition of a vector of time series, see [9]. In the first case, the advantage of a SEM is that the KF required to evaluate its likelihood is much simpler than that of a MEM, so the resulting procedure is faster and more stable; see Subsection 4.1 for further details. On the other hand, when the SEM form is used for signal extraction the estimated components have null variances and, therefore, do not change when the sample increases; see Subsection 4.2. This is a very important advantage, e.g., when the data is being adjusted for seasonality because it avoids the so called ‘revision effect’.

In this paper we revise and refine the discussion of these previous results and consider a new application, the extraction of a coincident trend indicator, in which MEM forms have substantial advantages. This happens because the components extracted from a MEM depend on both, past and future fluctuations, while those of a SEM collapse to exact values, depending only on the past. Because of this, the MEM trend includes future sample information that is not taken into account by the SEM trend. This result adds theoretical arguments to the practical benefits already pointed out by Proietti [23].

Therefore, our contribution with respect to previous works is twofold. First, we provide a systematic discussion of the strengths and weaknesses of both SS formulations for different applications. To this end, we draw on previous theoretical results and develop some new ones. These new results and the corresponding formal proofs are our second contribution.

The structure of the paper is as follows. Section 2 defines the MEM and SEM and presents the conditions required to assure their observational equivalence. Section 3 provides several propositions comparing the filtering and fixed-interval smoothing results obtained with both forms. Building on these results, Section 4 discusses the pros and cons of both representations for different uses, illustrating the discussion with practical examples. Finally, Section 5 provides some concluding remarks and suggests how to accommodate both forms in an eclectic ‘best of both worlds’ approach, which applies each representation to the task in which it has comparative advantage. The appendices provide proofs for the formal results.

2. Model formulations and equivalence

2.1. The Multiple Error Model

Consider the MEM model given by:

$$\mathbf{x}_{t+1} = \mathbf{\Phi} \mathbf{x}_t + \mathbf{E} \mathbf{w}_t \quad (2.1)$$

$$\mathbf{z}_t = \mathbf{H} \mathbf{x}_t + \mathbf{C} \mathbf{v}_t \quad (2.2)$$

where $\mathbf{z}_t \in \mathbb{R}^m$ is a random vector of endogenous variables or outputs, $\mathbf{x}_t \in \mathbb{R}^n$ is vector of state variables and \mathbf{w}_t , \mathbf{v}_t are conformable vectors of zero-mean white noise errors, such that:

$$\mathbb{E} \begin{bmatrix} \mathbf{w}_t \\ \mathbf{v}_t \end{bmatrix} \begin{bmatrix} \mathbf{w}_t & \mathbf{v}_t \end{bmatrix} = \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{R} \end{bmatrix} \quad (2.3)$$

In (2.1)-(2.3), $\mathbf{\Phi}$, \mathbf{E} , \mathbf{H} , \mathbf{C} , \mathbf{Q} , \mathbf{R} and \mathbf{S} are time-invariant coefficient matrices, being \mathbf{Q} and \mathbf{R} positive semi-definite. Last, the initial state is in general an unknown value: $\mathbf{x}_1 \sim (\bar{\mathbf{x}}_1, \mathbf{P}_1)$

Note that:

- 1) The state equation (2.1) characterizes the system dynamics while the observation equation (2.2) describes \mathbf{z}_t as the sum of linear combinations of the dynamic components, $\mathbf{H} \mathbf{x}_t$, and the observation errors, $\mathbf{C} \mathbf{v}_t$.
- 2) When we say that this is a MEM we refer to the fact that there are different errors affecting the state and observation equations, \mathbf{w}_t and \mathbf{v}_t , respectively.

3) Model (2.1)-(2.2) has no exogenous inputs. This assumption is made for simplicity, as the results in the following sections can be immediately extended to models with inputs.

The structure of the SS model makes it very adequate to estimate interesting signals buried in time series, such as seasonality, trends or cycles. In a SS framework, this ‘signal extraction’ is implemented by: (a) breaking up the state equation into the sub-systems characterizing each signal, (b) estimating the corresponding states and then, (c) estimating the relevant signals by combining the state estimates with chosen parts of the observation matrix \mathbf{H} , see e.g. [9]. Hence, the quality of signal extraction depends crucially on the properties of the state estimates, which will be discussed in Sections 3 and 4.

2.2. The Single Error Model

On the other hand, consider the SEM:

$$\mathbf{x}_{t+1}^* = \mathbf{\Phi} \mathbf{x}_t^* + \mathbf{K} \mathbf{a}_t \quad (2.4)$$

$$\mathbf{z}_t = \mathbf{H} \mathbf{x}_t^* + \mathbf{a}_t \quad (2.5)$$

where $\mathbf{x}_1^* \sim (\bar{\mathbf{x}}_1^*, \mathbf{P}_1^*)$ and the error term $\mathbf{a}_t \in \mathbb{R}^m$ is such that $\mathbf{a}_t \sim iid(\mathbf{0}, \mathbf{B})$. Note that the output \mathbf{z}_t and the matrices $\mathbf{\Phi}$ and \mathbf{H} are the same in both models, while the state vectors, the disturbances and error coefficient matrices are different in general.

2.3. Conditions for the equivalence of MEM and SEM models

At first sight, the SEM (2.4)-(2.5) seems to be a constrained version of the MEM (2.1)-(2.2). In fact, the relationships between both representations are more complex and have been discussed by the literature in several works.

Anderson and Moore [1] consider a time-varying parameter MEM and write then the corresponding filtering equations, which provide a time-varying SEM, which they call ‘innovations representation with finite initial time’. Assuming that the system output is stationary, they write the steady-state of these equations which coincides with (2.4)-(2.5). They call this model a ‘steady-state innovations form’.

Hannan and Deistler [17, Chapter 1] build on model (2.1)-(2.2) and then transform it into an equivalent (2.4)-(2.5) form assuming stability, i.e., that the eigenvalues of the transition matrix Φ lie within the unit circle. The error term in this model coincides with the one-step-ahead forecast error and, for this reason, they describe this model as ‘prediction error form’.

Aoki and Havenner [3] propose a model specification method that benefits from the SEM model structure and describe a procedure to obtain this representation when the error distribution matrices in (2.1)-(2.2), E and C , are constrained to identity.

Last Casals, Sotoca and Jerez [6] proved that z_t in (2.1)-(2.2) is also the output of (2.4)-(2.5), where the matrices K and B result from the equations:

$$K = (\Phi P H^T + E S C^T) B^{-1} \quad (2.6)$$

$$B = H P H^T + C R C^T \quad (2.7)$$

...and P is set to the unique strong solution of the algebraic Riccati equation:

$$P = \Phi P \Phi^T + E Q E^T - K B K^T \quad (2.8)$$

This result requires two unrestrictive assumptions about (2.1)-(2.2). First, the ‘time immemorial hypothesis’ [11] implying that the system started in an infinitely remote past. Second, that the model is detectable, so any non-observable mode corresponds to an eigenvalue of Φ within the unit circle [1, Chapter 4] being this condition necessary and sufficient to assure that (2.8) has a unique strong solution, [13, Theorem 3.2].

Note that the matrix B in Eq. (2.6) is to be inverted. If the system output includes deterministic terms, the determinant of this matrix is zero and the SS model would be an “improper formulation.” In this case there are two possible solutions, either transforming the formulation into a proper equivalent form, or using the Moore-Penrose pseudo-inverse of B . Whenever possible, the first treatment should be preferred, as it will result in a faster and more stable performance. The same considerations apply to matrix B_t in Eq. (3.3).

Under these conditions, the matrices K , B and P resulting from (2.6)-(2.8) characterize a SEM observationally equivalent to the MEM (2.1)-(2.2) because both have the same likelihood

[6, Appendix]. As we will see in next Section, these matrices are steady-state solutions for the KF of model (2.1)-(2.2). Equations (4.1)-(4.2) and (4.3)-(4.4) in Section 4 provide an example of two equivalent MEM and SEM.

3. Signal extraction for MEM and SEM

3.1. Kalman filtering

The KF for the MEM (2.1)-(2.2) computes one-step-ahead estimates of the states as well as the corresponding error covariance matrix by propagating the following equations, [24]:

$$\mathbf{x}_{t+1|t} = \Phi \mathbf{x}_{t|t-1} + \mathbf{K}_t \tilde{\mathbf{z}}_{t|t-1} \quad (3.1)$$

$$\tilde{\mathbf{z}}_{t|t-1} = \mathbf{z}_t - \mathbf{H} \mathbf{x}_{t|t-1} \quad (3.2)$$

$$\mathbf{K}_t = (\Phi \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{E} \mathbf{S} \mathbf{C}^T) (\mathbf{B}_t)^{-1} \quad (3.3)$$

$$\mathbf{B}_t = \mathbf{H} \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{C} \mathbf{R} \mathbf{C}^T \quad (3.4)$$

$$\mathbf{P}_{t+1|t} = \Phi \mathbf{P}_{t|t-1} \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T - \mathbf{K}_t \mathbf{B}_t (\mathbf{K}_t)^T \quad (3.5)$$

where $\mathbf{x}_{t+1|t}$ and $\mathbf{x}_{t|t-1}$ denote respectively the expectation of the states at time $t+1$ and t , conditional on the information available up to the previous period. We will often refer to these moments as ‘Kalman states’. On the other hand, $\mathbf{P}_{t+1|t}$ and $\mathbf{P}_{t|t-1}$ are the conditional covariances corresponding to the Kalman states.

Note that: (a) starting this recursion requires suitable initial conditions, $\bar{\mathbf{x}}_1$ and \mathbf{P}_1 , see e.g., [11] or [8]; (b) the SEM (2.4)-(2.5) is similar to (3.1)-(3.2) after reordering the latter expression to $\mathbf{z}_t = \mathbf{H} \mathbf{x}_{t|t-1} + \tilde{\mathbf{z}}_{t|t-1}$; and (c) equations (2.6), (2.7) and (2.8) are the steady-state versions of (3.3), (3.4) and (3.5), respectively. Therefore, the SEM matrices can be derived using the steady-state values of a KF.

On the other hand, if we enforce observational equivalence between models (2.1)-(2.2) and (2.4)-(2.5), any adequately initialized KF will be such that:

$$\mathbf{B}_t = \mathbf{B}_t^* \quad (3.6)$$

$$\mathbf{K}_t = \mathbf{K}_t^* \quad (3.7)$$

$$\tilde{\mathbf{z}}_{t|t-1} = \tilde{\mathbf{z}}_{t|t-1}^* \quad (3.8)$$

$$\mathbf{x}_{t|t-1} = \mathbf{x}_{t|t-1}^* \quad (3.9)$$

where the ‘starred’ values correspond to the SEM (2.4)-(2.5). The intuition behind these results is that, if two models are observationally equivalent, all the filtering results that affect the likelihood value must coincide. As a consequence, both forms yield the same one-step-ahead forecasts and Kalman states for all t . On the other hand their state covariances are not the same. Their differences are formally described in the following results.

Proposition 1: Under the immemorial time hypothesis [11] the covariances of the initial states in the MEM and SEM are such that:

1.a: If the system is stationary:

$$\text{cov}(\mathbf{x}_1) = \text{cov}(\mathbf{x}_1^*) + \mathbf{P} \quad (3.10)$$

1.b: If the system is nonstationary, but detectable:

$$\text{cov}[\mathbf{x}_{T+1/T}] = \text{cov}[\mathbf{x}_{T+1/T}^*] + \mathbf{P} \quad (3.11)$$

where \mathbf{P} is the strong solution of the algebraic Riccati equation, defined in (2.8) and T is the minimal sample size required to achieve a finite $\text{cov}[\mathbf{x}_{T+1/T}]$ starting from diffuse initial conditions.

Proof. See Appendix A.1.

Proposition 2. The *a priori* and *a posteriori* filtering covariances in the Riccati equation (3.5) are such that, if $\mathbf{P}_{t|t-1} = \mathbf{P}_{t|t-1}^* + \mathbf{P}$, then:

$$\mathbf{P}_{t+1|t} = \mathbf{P}_{t+1|t}^* + \mathbf{P} \quad (3.12)$$

Proof. See Appendix A.2. This proposition also holds for samples with missing data, see Appendix B.

Proposition 3. Under detectability the covariances of the filtered states in the SEM and MEM are such that:

$$\lim_{t \rightarrow +\infty} \mathbf{P}_{t|t-1}^* = \mathbf{0} ; \lim_{t \rightarrow +\infty} \mathbf{P}_{t+1|t}^* = \mathbf{0} \quad (3.13)$$

$$\lim_{t \rightarrow +\infty} \mathbf{P}_{t|t-1} = \mathbf{P} ; \lim_{t \rightarrow +\infty} \mathbf{P}_{t+1|t} = \mathbf{P} \quad (3.14)$$

Proof. This is a known result which can be proved building on the convergence properties of the algebraic Riccati equation to its strong solution; see [14] and [22]. The speed of convergence to this solution is a quadratic function of the eigenvalues of $\Phi - \mathbf{K}\mathbf{H}$. Therefore, if they are strictly within the unit circle (which is assured if the model is invertible) the speed of convergence to the strong solution is exponential and, as a consequence, so is the speed of convergence in (3.13) and (3.14).

The implications of these results are clear. Proposition 1 states that the uncertainty about the initial state of the SEM is less or equal than that of the MEM, while Proposition 2 assures that this inequality is maintained throughout all the subsequent recursions of the KF. On the other hand, the limit values given by Proposition 3 assure that: (a) the conditional state values derived by any signal-extraction procedure applied to the SEM converge to values with no uncertainty and, in the limit, to the ‘true’ states, \mathbf{x}_t^* ; while (b) in general, the state estimates corresponding to a MEM will remain uncertain asymptotically, see (3.14).

3.2. Fixed-interval smoothing

The smoothing equations for the MEM model are given by the following backward recursion [10]:

$$\mathbf{x}_{t|N} = \mathbf{x}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{r}_{t-1} \quad (3.15)$$

$$\mathbf{P}_{t|N} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{R}_{t-1} \mathbf{P}_{t|t-1} \quad (3.16)$$

$$\mathbf{r}_{t-1} = \bar{\Phi}_t^T \mathbf{r}_t + \mathbf{H}^T (\mathbf{B}_t)^{-1} \tilde{\mathbf{z}}_{t|t-1} \quad \text{with } \mathbf{r}_N = \mathbf{0} \quad (3.17)$$

$$\mathbf{R}_{t-1} = \mathbf{H}^T (\mathbf{B}_t)^{-1} \mathbf{H} + \bar{\Phi}_t^T \mathbf{R}_t \bar{\Phi}_t \quad \text{with } \mathbf{R}_N = \mathbf{0} \quad (3.18)$$

$$\bar{\Phi}_t = \Phi - \mathbf{K}_t \mathbf{H} \quad (3.19)$$

$$\mathbf{B}_t = \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T + \mathbf{C}\mathbf{R}\mathbf{C}^T \quad (3.20)$$

where $\mathbf{x}_{t|N}$ is the expectation of the states at time t , conditional on all the information available on a sample of size N and $\mathbf{P}_{t|N}$ is the corresponding conditional covariance. We will refer to these conditional expectations as ‘smoothed states.’

Writing the analogous equations for SEM model is easy but not very useful. Note that: (a) according to (3.14) the state covariance $\mathbf{P}_{t|t-1}^*$ converges to zero as t increases, and (b) the difference between the filtered and smoothed states depends on this covariance, see (3.15). As a consequence both, the smoothed and filtered states for the SEM model ($\mathbf{x}_{t|N}^*$ and $\mathbf{x}_{t|t-1}^*$) will converge to the true states (\mathbf{x}_t^*) as t increases. It can be shown, see [9, Section 4], that the covariance of the smoothed states at time t is $\mathbf{P}_{t|N} = (\mathbf{\Phi} - \mathbf{K}\mathbf{H}) \mathbf{P}_{t|N} (\mathbf{\Phi} - \mathbf{K}\mathbf{H})^T$, so the speed of this convergence of the uncertainty to zero is quadratic and governed again by the eigenvalues of $\mathbf{\Phi} - \mathbf{K}\mathbf{H}$.

Figure 3.1 illustrates the convergence of the trace of the smoothing covariance, $\text{tr}(\mathbf{P}_{t|N})$, for different model structures. The calculations have been done using 50 simulated samples of the data generating process indicated in each case. We did not compute different realizations for each process because the convergence path is governed by its MA roots, so the trace paths corresponding to different samples of the same process would be identical.

[Insert Figure 3.1]

Panel (a) shows the result corresponding to an ARIMA(2,1,0) model. In this case, where there are no MA roots, the convergence to zero occurs after processing a number of samples equal to the number of AR roots, which is three in this case. Note that the presence of a unit AR root does not affect the convergence. Panels (b), (c) and (d) show the analogous trace profile corresponding to an ARIMA(2,1,1) process. Note that, as the MA term approaches noninvertibility, the trace takes longer to converge. In the last case, when the model is noninvertible, the value of the trace on convergence is 1/1000.

Proposition 4: If two MEM and SEM models are observationally equivalent, the differences between their smoothed states and their covariances are given by:

$$\mathbf{x}_{t|N} - \mathbf{x}_{t|N}^* = \mathbf{P}\mathbf{r}_{t-1} \quad (3.21)$$

$$\mathbf{P}_{t|N} - \mathbf{P}_{t|N}^* = \mathbf{P} - \mathbf{P} \mathbf{R}_{t-1} \mathbf{P} - \mathbf{P}_{t|t-1}^* \mathbf{R}_{t-1} \mathbf{P} - \mathbf{P} \mathbf{R}_{t-1} \mathbf{P}_{t|t-1}^* \quad (3.22)$$

where the ‘star’ symbol denotes again the results of the SEM, \mathbf{P} is the strong solution of (2.8) and the sequences \mathbf{r}_{t-1} and \mathbf{R}_{t-1} were defined in (3.17)-(3.18) respectively.

Proof: As we saw in Sub-section 3.1, observational equivalence between a MEM and a SEM implies that many filtering results coincide, so $\mathbf{x}_{t|t-1} = \mathbf{x}_{t|t-1}^*$, $\tilde{\mathbf{z}}_{t|t-1} = \tilde{\mathbf{z}}_{t|t-1}^*$, $\mathbf{K}_t = \mathbf{K}_t^*$, $\mathbf{B}_t = \mathbf{B}_t^*$ and, as a consequence, $\bar{\Phi}_t = \bar{\Phi}_t^*$, $\mathbf{r}_t = \mathbf{r}_t^*$ and $\mathbf{R}_t = \mathbf{R}_t^*$. Substituting these equalities and $\mathbf{P}_{t|t-1} = \mathbf{P}_{t|t-1}^* + \mathbf{P}$ in (3.15)-(3.20) yields immediately (3.21) and (3.22). This result also holds for samples with missing data, see Appendix B.

Proposition 5. The covariances of the smoothed states in a SEM and a MEM are such that:

$$\lim_{t \rightarrow +\infty} \mathbf{P}_{t|N}^* = \mathbf{0} \quad (3.23)$$

$$\lim_{t \rightarrow +\infty} \mathbf{P}_{t|N} = \mathbf{P} - \mathbf{P} \mathbf{R}_{t-1} \mathbf{P} \quad (3.24)$$

Proof: Applying Proposition 3 to (3.16) immediately implies that the ‘starred’ covariances tend to zero as the sample increases, proving (3.23). Doing the same with expression (3.22) implies, accordingly, expression (3.24).

These results have consequences of practical interest:

- 1) As the smoothed and filtered states of the SEM, $\mathbf{x}_{t|N}^*$ and $\mathbf{x}_{t|t-1}^*$, converge to the true states, \mathbf{x}_t^* , the dynamic components driving the system will be estimated with null uncertainty.
- 2) According to (3.21), the smoothed states for the MEM are, $\mathbf{x}_{t|N} = \mathbf{x}_{t|N}^* + \mathbf{P} \mathbf{r}_{t-1}$, so they combine the dynamics of $\mathbf{x}_{t|N}^*$ with those of the addend $\mathbf{P} \mathbf{r}_{t-1}$.
- 3) The dynamics of \mathbf{r}_{t-1} are given by the transition matrix $\bar{\Phi}_t = \Phi - \mathbf{K}_t \mathbf{H}$ which, if the system is invertible converges to $\Phi - \mathbf{K} \mathbf{H}$ as t increases. Note that this matrix characterizes the moving average structure in the SEM (2.4)-(2.5).
- 4) \mathbf{r}_{t-1} is excited by the standardized future innovations through the term $\mathbf{H}^\top (\mathbf{B}_t)^{-1} \tilde{\mathbf{z}}_{t|t-1}$, so the variables $\mathbf{x}_{t|t-1}$ and \mathbf{r}_{t-1} are independent.

The previous analysis is based on standard filtering and smoothing recursions, under the assumption that they were started with adequate initial conditions $\bar{\mathbf{x}}_1$ and \mathbf{P}_1 . In practice this requires assuming either: (a) that the system to be stationary, which is the only case where there is a finite closed form for $\bar{\mathbf{x}}_1$ and \mathbf{P}_1 , or (b) that a specialized procedure has been applied to an initial stretch of data to isolate the effect of diffuse initial conditions, see [7]. The last idea is very important, as it allows us to extend previous results to the nonstationary case.

4. Pros and cons of SEM and MEM

The results in previous section provide three main conclusions. First, the uncertainty of the filtered and smoothed states of the SEM converges to zero, while the corresponding MEM states remain uncertain. Second, current smoothed states in the MEM are affected by past and future values of the output variable, while those in the equivalent SEM are only affected by the past. Third, the dynamics of the smoothed states for the SEM are determined by dynamics of the system, while those of the MEM combine the system dynamics with an additional forward term, governed by the matrix $\Phi - \mathbf{K}_t \mathbf{H}$, which converges to $\Phi - \mathbf{K} \mathbf{H}$ as \mathbf{K}_t converges to the steady-state gain of the KF, \mathbf{K} .

These properties determine the pros and cons of MEM and MEM representations in three areas: likelihood computation, seasonal adjustment and extraction of a trend indicator.

4.1. Likelihood computation

Computing the Gaussian likelihood in prediction error decomposition form requires estimating the model innovations, see Terceiro [24]. In a SS framework, this is typically done using the KF. Using the equivalent SEM to do this has important advantages arising from the fact that, in this representation, the null matrix is an exact and strong solution of the algebraic Riccati equation of the filter. Taking advantage of this property one can devise a faster and more stable specialized filter, see [6] and [12].

Superior computational speed results from a positive tradeoff between the overhead required to obtain the SEM representation and the computational savings that the simplified filter provides. Obviously, these gains depend on the number of endogenous variables and the dimension of the state vector. In the cases considered in [6], the ratio between the number of

basic floating-point operations required by a standard algorithm and the improved one ranges from 1.26 to 3.61. For models with seasonal structure this ratio is one order of magnitude larger or best.

On the other hand, the algorithm based on the SEM model is numerically stable because it avoids the propagation of the Riccati equation which the literature, see e.g. [4], unanimously recognizes as the main source of numerical degradation of the KF.

4.2. Seasonal adjustment

Working with a SEM form has also substantial advantages when one wants to decompose a series into its trend, cycle, seasonal and irregular components, in particular when the final goal of the exercise is to adjust for seasonality. Casals, Jerez and Sotoca [9] present a structural decomposition method building on this idea. Here we will illustrate these advantages with a simulated dataset that was already employed in [8, Section 6].

Consider the data generating process displayed in Table 4.1, where the trend is an I(2) process and the seasonal component is a quarterly dummy variable model [15] such that the sum of the seasonal components over the year is white noise.

[Insert Table 4.1]

To simplify the comparison with the Hodrick and Prescott [18] filter, we set the ratio between the variances of the irregular component and the trend to the value assumed by these authors for quarterly data. This is therefore the SS model implied by this filter [16] with an additional component to allow for seasonality.

The model in Table 4.1 can be easily written in MEM form. Table 4.2 displays the equivalent SEM.

[Insert Table 4.2]

We generated 200 sample draws from this process and extracted the structural components from both representations using the fixed-interval smoothing algorithm described in [7]. Figure 4.1.a compares the traces of the corresponding smoother covariance matrices, $\mathbf{P}_{t|N}$ and $\mathbf{P}_{t|N}^*$.

Figure 4.1.b displays the same trace sequence but computed for a sample with missing values in observations 100 to 104.

[Insert Figures 4.1.a and 4.1.b]

Note that:

- 1) The uncertainty of the MEM states displays the ‘U-shape’ characteristic of symmetric filters, so the estimates at the center of the sample are more precise than those at the extremes.
- 2) This uncertainty profile implies that MEM components will suffer ‘revisions’, i.e., augmenting the sample generates a change in the smoothed estimate for each component which is proportional to its uncertainty.
- 3) On the other hand, as stated by Proposition 5, the variance of the components resulting from the SEM model converges to exact values with null variances and, therefore, will not be affected by new data.
- 4) The case of missing data is of some interest, either because the sample actually includes missing values, or because this is an effective way to treat the presence of impulse-type outliers. Figure 4.1.b shows that missing in-sample values create a transient peak in the uncertainty of the states that does not alter substantially the uncertainty profile shown in Figure 4.1.a. Appendix B proves the validity of propositions 2 and 4 in the case of missing values.

Besides the lack of revisions, the SEM has other advantages for seasonal adjustment. As stated in Section 3, the smoothed states of the SEM only depend on the system dynamics, so the dynamic structure of its components should match exactly with theory. On the other hand, the smoothed states of the MEM model combine the system dynamics with a forward moving average structure governed by the matrix $\Phi - \mathbf{K}_t \mathbf{H}$, so some distortion in the dynamics of its components is to be expected.

To illustrate this distortion, Figures 4.2.a and 4.2.b display the sample autocorrelation function of the four-quarters rolling sum of the seasonal component computed using the MEM and SEM forms, respectively. According to the theoretical model for the seasonal component,

the first series should be white noise while the second should follow a long memory moving average process. In both cases there should be no seasonal autocorrelations.

[Insert Figures 4.2.a and 4.2.b]

Figure 4.2.a displays strong regular and seasonal autocorrelation patterns, drastically different from its theoretical white noise structure. This distortion is due to the forward structure characterized in Proposition 4. On the other hand, the pattern in Figure 4.2.b fits well with the theoretical structure of the SEM component, which is a long moving average. In particular, it does not show significant autocorrelations in the seasonal lags 4, 8 and 12.

Therefore we conclude that SEM representation has two advantages for seasonal adjustment, as the corresponding adjusted series: (a) do not change as the sample increases, and (b) follow strictly the system dynamics. In comparison, the series resulting from the equivalent MEM are affected by revisions and contaminated by spurious regular and seasonal dynamics. As we will see in next sub-section, this conclusion is reversed when one wants to compute a trend indicator.

4.3. Trend extraction

Now we will discuss the pros and cons of MEM and SEM for trend extraction using a time series of US Gross Domestic Product (GDP), from 1952-Q1 to 2012-Q3, in billions of chained 2005 dollars, which was obtained from the US Bureau of Economic Analysis (www.bea.gov).

We first fitted an $I(2) + \text{error}$ model to this data, with the following results:

$$\begin{bmatrix} \mu_{t+1} \\ \beta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \zeta_t$$

$$GDP_t = \mu_t + \hat{\varepsilon}_t \quad (4.1)$$

$$\text{cov} \begin{bmatrix} \zeta_t \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} 1.641 \times 10^4 / 1600 & 0 \\ 0 & 1.641 \times 10^4 \end{bmatrix} \quad (4.2)$$

...where GDP_t denotes the GDP at quarter t , μ_t is the latent trend component and β_t is the change in the trend from time $t-1$ to t . The estimates have been computed by Gaussian maximum likelihood and the noise-to-variance ratio has been once again constrained to the $1/1600$ value

assumed by Hodrick and Prescott [18] for quarterly data. Then, we solved equations (2.6)-(2.8) with the parameters in (4.1)-(4.2) to obtain the equivalent SEM:

$$\begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + \begin{bmatrix} 0.223 \\ 0.0224 \end{bmatrix} \hat{a}_t \quad (4.3)$$

$$GDP_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + \hat{a}_t ; \hat{\sigma}_a^2 = 2.052 \times 10^4 \quad (4.4)$$

Last, we applied the smoothing algorithm in [8] to both models to obtain estimates, $\hat{\mu}_{t|N}$ and $\hat{x}_{t|N}^1$, for the trend component of the MEM and SEM, respectively. Figure 4.3 shows the profile of the first-order difference of the trend obtained with both models.

[Insert Figure 4.3]

Note that changes in the trend obtained from the MEM model are much smoother. This happens because this model has two errors, so the trend error variance can be set to a small value relative to the variance of the observation error, as we have done in this case. Such a control on the noise-variance-ratio is not possible in the SEM model, which has a single error affecting both equations. The capacity to control this ratio is convenient for some uses, e.g. empirical macroeconomics, as it allows one to filter out short-term fluctuations and focus accordingly on the long-term behavior of the time series analyzed.

Figure 4.3 also suggests that there is a shift in phase between both series, so ups and downs in the MEM trend lead those of the SEM trend. To test and measure this feature, we fitted dynamic regressions between both trend indicators with the following results:

$$\nabla^2 \hat{x}_{t|N}^1 = .185 - 94.242 \nabla^2 \hat{\mu}_{t|N} + 167.918 \nabla^2 \hat{\mu}_{t-1|N} - 73.468 \nabla^2 \hat{\mu}_{t-2|N} + \hat{u}_t^{SSE} ; R_{SSE}^2 = .404 \quad (4.5)$$

(.667) (12.079) (23.717) (12.092)

$$\nabla^2 \hat{\mu}_{t|N} = .059 + .026 \nabla^2 \hat{x}_{t|N}^1 + .010 \nabla^2 \hat{x}_{t-1|N}^1 - .001 \hat{x}_{t-2|N}^1 + \hat{u}_t^{MSE} ; R_{MSE}^2 = .059 \quad (4.6)$$

(.104) (.008) (.009) (.008)

where ∇^2 is the second-order difference operator and the figures in parentheses are the coefficient standard errors.

The coefficients and R-squared values in (4.5) indicate clearly that lagged values of the MEM trend, $\hat{\mu}_{t|N}$, affect current values of the SEM trend, while (4.6) shows that the opposite is

not true. This lag-lead structure is explained by the fact that the MEM trend at time t is affected by future sample values $GDP_{t+1}, GDP_{t+2}, \dots, GDP_{t+N}$; see the discussion of Proposition 5. This does not happen with the smoothed states resulting from the SEM. According to Eq. (3.15), when the filter covariance $P_{t|t-1}$ is null, the smoothed states are not affected by the future information contained in the term r_{t-1} . Because of this, the MEM trend predicts future fluctuations of the SEM trend and, therefore, it is more useful as trend indicator.

On the other hand, these indicators are typically extracted using seasonally adjusted data so, in this case, the best approach would probably consist in using the SEM to obtain the adjusted series, and then computing the trend indicator using the corresponding ‘seasonally free’ MEM.

5. Concluding remarks

The discussion in Section 4 suggests that MEM and SEM forms can be combined in a ‘best of both worlds’ approach.

A SEM has advantages in terms of speed and stability when one wants to compute the likelihood of a fixed-coefficients SS model. It also provides a structural decomposition that follows strictly the system dynamics and avoids revisions. Therefore, it is better suited than the equivalent MEM when one wants to remove a given component, e.g., seasonality, from the series.

On the other hand, a MEM has two advantages to compute a trend indicator. First, each value in the trend component combines both, past and future information, so current values of the MEM trend lead the SEM trend. Second, a representation with multiple errors allows one to impose noise variance ratio constraints and, therefore, to choose the ‘smoothness’ of the trend component, being this feature very convenient to detect business cycle turning points.

As a consequence of the previous discussion, our point of view about the formulation of choice is rather eclectic, and can be summarized in the following points:

- 1) For computational purposes, it is convenient in general to use a SS representation because it allows one to apply the same procedures to the wide family of standard formulations that can be written as an equivalent SS form, such as VARIMA or

structural equation models, as well as to the models implied by popular filters, such as those of Hodrick-Prescott, [18], or Beveridge-Nelson, [21].

- 2) To compute the Gaussian likelihood of a fixed-coefficient model in prediction-error form, [24], the SEM representation is always superior, as it produces the same likelihood value with more stability and less computational cost. Accordingly, we think that this is an aspect that should be automatically managed by the software, without requiring user intervention.
- 3) About the last applications, decomposing a time series without revisions or computing a coincident trend indicator, user intervention is critical. This happens because the underlying computational process is the same in both cases and, therefore, choosing the right representation requires the insight of the analyst. Our recommendation is using a SEM when one wants to avoid revisions and a MEM when taking advantage of future information is important, as happens in macroeconomic analyses.

The procedures described in this article are implemented in the E⁴ functions ‘sstoinn’ (conversion of a MEM to the equivalent SEM), ‘lffast’ (likelihood computation with automatic conversion to the SEM form) and ‘e4trend’ (structural decomposition with optional conversion to the SEM form). E⁴ is a MATLAB toolbox for time series modeling, which can be downloaded at: www.ucm.es/info/icae/e4. The source code for all the functions in the toolbox is freely provided under the terms of the GNU General Public License. This site also includes a complete user manual and other reference materials.

Acknowledgements: We benefitted from the ideas and suggestions by Antonio Garcia-Ferrer, Marcos Bujosa, Esther Ruiz and an anonymous referee. Their contribution is gratefully acknowledged although any errors are our sole responsibility. This work was finished while Sonia Sotoca and Miguel Jerez were visiting UC Davis. Special thanks are due to the Department of Economics for providing ideal working conditions. Financial support from the *Programa Becas Complutenses Del Amo*, from Universidad Complutense de Madrid, UC Davis and *Ministerio de Economia y Competitividad*, through Grant ECO2011-23972, is also gratefully acknowledged.

Appendix A. Proofs for Propositions 1 and 2

A.1. Proof of Proposition 1

Let us consider first the stationary case. Denoting $\text{cov}(\mathbf{x}_1) = \mathbf{P}_1$ and $\text{cov}(\mathbf{x}_1^*) = \mathbf{P}_1^*$ the state equations for the MEM and SEM forms imply, respectively:

$$\mathbf{P}_1 = \Phi \mathbf{P}_1 \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T \quad (\text{A.1})$$

$$\mathbf{P}_1^* = \Phi \mathbf{P}_1^* \Phi^T + \mathbf{K} \mathbf{B} \mathbf{K}^T \quad (\text{A.2})$$

so, subtracting (A.1) and (A.2) yields:

$$\mathbf{P}_1 - \mathbf{P}_1^* = \Phi (\mathbf{P}_1 - \mathbf{P}_1^*) \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T - \mathbf{K} \mathbf{B} \mathbf{K}^T \quad (\text{A.3})$$

Therefore, $\mathbf{P} = \mathbf{P}_1 - \mathbf{P}_1^*$ is the strong solution of the Riccati equation (A.3), being $\mathbf{K} = (\Phi \mathbf{P} \mathbf{H}^T + \mathbf{E} \mathbf{S} \mathbf{C}) \mathbf{B}^{-1}$. Accordingly, $\mathbf{P} = \mathbf{P}_1 - \mathbf{P}_1^*$ is positive semi-definite. ■

In the nonstationary case, the initial state \mathbf{x}_1 has diffuse components, so $\mathbf{P}_1 \rightarrow \infty$. The immemorial time assumption [11] allows one to decompose \mathbf{x}_1 into two terms which are orthogonal by construction:

$$\mathbf{x}_1 = \mathbf{x}_{1/0,-1,\dots} + \tilde{\mathbf{x}}_1 \quad (\text{A.4})$$

where $\mathbf{x}_{1/0,-1,\dots} = E[\mathbf{x}_1 / \mathbf{z}_0, \mathbf{z}_{-1}, \dots]$ is the nonstationary part of \mathbf{x}_1 and $\tilde{\mathbf{x}}_1$ is the stationary component. As detectability is a necessary and sufficient condition for the convergence of the Riccati equation, see [14] and [22], we have: $E[\tilde{\mathbf{x}}_1] = \mathbf{0}$ and $\text{cov}[\tilde{\mathbf{x}}_1] = \mathbf{P} = \Phi \mathbf{P} \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T - \mathbf{K} \mathbf{B} \mathbf{K}^T$

To cope with this situation, Casals, Jerez and Sotoca [7] proposed using the initial conditions for the stationary case and then correcting their effect through the following equations:

$$\mathbf{x}_{t+1|t} = \mathbf{x}_{t+1|t}^E + \bar{\bar{\Phi}}_{t+1} \mathbf{x}_{1|t} \quad (\text{A.5})$$

$$\mathbf{P}_{t+1|t} = \mathbf{P}_{t+1|t}^E + \bar{\bar{\Phi}}_{t+1} \mathbf{P}_{1|t} (\bar{\bar{\Phi}}_{t+1})^T \quad (\text{A.6})$$

$$\mathbf{x}_{1|t} = \mathbf{P}_{1|t} (\mathbf{P}_1^{-1} \bar{\mathbf{x}}_1 + \mathbf{w}_t) \quad (\text{A.7})$$

$$\mathbf{P}_{1|t} = (\mathbf{P}_1^{-1} + \mathbf{W}_t)^{-1} \quad (\text{A.8})$$

where:

$$\bar{\bar{\Phi}}_{t+1} = (\Phi - \mathbf{K}_t \mathbf{H}) \bar{\bar{\Phi}}_t \text{ con } \bar{\bar{\Phi}}_1 = \mathbf{I} \quad (\text{A.9})$$

$$\mathbf{W}_t = \mathbf{W}_{t-1} + (\mathbf{X}_t)^\top \mathbf{B}_t^{-1} \mathbf{X}_t \text{ con } \mathbf{W}_0 = \mathbf{0} \quad (\text{A.10})$$

$$\mathbf{w}_t = \mathbf{w}_{t-1} + (\mathbf{X}_t)^\top \mathbf{B}_t^{-1} \tilde{\mathbf{z}}_t \text{ con } \mathbf{w}_0 = \mathbf{0} \quad (\text{A.11})$$

$$\mathbf{X}_t = \mathbf{H} \bar{\bar{\Phi}}_t \quad (\text{A.12})$$

and $\mathbf{x}_{t+1|t}^E$, $\mathbf{P}_{t+1|t}^E$, \mathbf{K}_t , \mathbf{B}_t , and $\tilde{\mathbf{z}}_t$ result from the propagation of a KF with initial conditions $(\mathbf{0}, \mathbf{P})$, thus ignoring the diffuse component of \mathbf{x}_1 . The terms $\bar{\bar{\Phi}}_{t+1} \mathbf{x}_{1|t}$ and $\bar{\bar{\Phi}}_{t+1} \mathbf{P}_{1|t} (\bar{\bar{\Phi}}_{t+1})^\top$ correct the arbitrary filter initialization, taking into account the diffuse component of \mathbf{x}_1 . After processing a sufficient number of observations T , detectability assures that the corrected filter given by (A.5)-(A.12) collapses to the standard KF (3.1)-(3.5).

If the model has no inputs, $\bar{\mathbf{x}}_1 = \mathbf{0}$ in (A.6)-(A.7) and, therefore, \mathbf{P}_1^{-1} can be computed following [11]. In the SEM the procedure is further simplified as the terms and $\mathbf{x}_{t+1|t}^E$, $\mathbf{P}_{t+1|t}^E$, \mathbf{K}_t , \mathbf{B}_t , and $\tilde{\mathbf{z}}_t$ in (A.5)-(A.12) result from a KF initialized with $(\mathbf{0}, \mathbf{0})$, because the strong solution to the algebraic Riccati equation for the SEM is null.

It is immediate to see that the correction terms $\bar{\bar{\Phi}}_{t+1} \mathbf{x}_{1|t}$ and $\bar{\bar{\Phi}}_{t+1} \mathbf{P}_{1|t} (\bar{\bar{\Phi}}_{t+1})^\top$ in (A.5)-(A.6) are identical for the equivalent MEM and SEM, as its computation only depends on \mathbf{K}_t and \mathbf{B}_t , that coincide for both forms.

Then, taking into account that $\mathbf{P}_1^E = \mathbf{P}$ in the MEM and $\mathbf{P}_1^E = \mathbf{0}$ in the SEM, Proposition 2 implies that $\mathbf{P}_{t+1|t}^E = \mathbf{P}_{t+1|t}^{*E}$ for all t and, after processing the corrected filter given by (A.5)-(A.12), we have that:

$$\mathbf{P}_{T+1|T} = \mathbf{P}_{T+1|T}^* + \mathbf{P} \quad (\text{A.13})$$

where the ‘star’ denotes that these variables correspond to the SEM. ■

A.2. Proof of Proposition 2

The Riccati equations of the KF for the MEM and SEM forms are, respectively:

$$\mathbf{P}_{t+1|t} = \Phi \mathbf{P}_{t|t-1} \Phi^\top + \mathbf{E} \mathbf{Q} \mathbf{E}^\top - \mathbf{K}_t \mathbf{B}_t (\mathbf{K}_t)^\top \quad (\text{A.14})$$

$$\mathbf{P}_{t+1|t}^* = \Phi \mathbf{P}_{t|t-1}^* \Phi^\top + \mathbf{K} \mathbf{B} \mathbf{K}^\top - \mathbf{K}_t^* \mathbf{B}_t^* (\mathbf{K}_t^*)^\top \quad (\text{A.15})$$

so their difference is:

$$\mathbf{P}_{t+1|t} - \mathbf{P}_{t+1|t}^* = \Phi (\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}^*) \Phi^\top + \mathbf{E} \mathbf{Q} \mathbf{E}^\top - \mathbf{K} \mathbf{B} \mathbf{K}^\top - \mathbf{K}_t \mathbf{B}_t (\mathbf{K}_t)^\top + \mathbf{K}_t^* \mathbf{B}_t^* (\mathbf{K}_t^*)^\top \quad (\text{A.16})$$

Observational equivalence between the MEM and SEM forms implies that $\mathbf{B}_t = \mathbf{B}_t^*$ and $\mathbf{K}_t = \mathbf{K}_t^*$, see (3.6) and (3.7). Therefore, we can cancel the last two addends in the right-hand-side of (A.16) and this expression simplifies to:

$$\mathbf{P}_{t+1|t} - \mathbf{P}_{t+1|t}^* = \Phi (\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}^*) \Phi^\top + \mathbf{E} \mathbf{Q} \mathbf{E}^\top - \mathbf{K} \mathbf{B} \mathbf{K}^\top \quad (\text{A.17})$$

Last, note that (A.17) is a Riccati equation in which the term in parentheses is, by hypothesis, $\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}^* = \mathbf{P}$, so the left-hand-side will be accordingly: $\mathbf{P}_{t+1|t} - \mathbf{P}_{t+1|t}^* = \mathbf{P}$ ■

Appendix B. Validity of propositions 2 and 4 for samples with missing data

Proposition 2 states that, if at time t $\mathbf{P}_{t+1|t} - \mathbf{P}_{t+1|t}^* = \mathbf{P}$, then this distance will be kept throughout the rest of the propagation. Now we will prove that this result holds even when the sample includes missing values.

With missing data, the observation equation of the MEM can be written as:

$$\mathbf{z}_t = \mathbf{A}_t (\mathbf{H} \mathbf{x}_t + \mathbf{C} \mathbf{v}_t) \quad (\text{B.1})$$

where \mathbf{A}_t is a binary time-varying matrix manifesting the observability of \mathbf{z}_t at time t , so a null value in this matrix indicates that the corresponding output value is missing, while a one corresponds to a non-missing value.

The Riccati equations in the KF for the MEM and SEM are, respectively:

$$\mathbf{P}_{t+1|t} = \Phi \mathbf{P}_{t|t-1} \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T - \mathbf{K}_t \mathbf{B}_t (\mathbf{K}_t)^T \quad (\text{B.2})$$

$$\mathbf{P}_{t+1|t}^* = \Phi \mathbf{P}_{t|t-1}^* \Phi^T + \mathbf{K} \mathbf{B} \mathbf{K}^T - \mathbf{K}_t^* \mathbf{B}_t^* (\mathbf{K}_t^*)^T \quad (\text{B.3})$$

where the starred values correspond again to the SEM. On the other hand, if the sample includes missing values it is straightforward to prove that $\mathbf{B}_t = \mathbf{B}_t^*$ and $\mathbf{K}_t \mathbf{B}_t = \mathbf{K}_t^* \mathbf{B}_t^*$. Taking into account that $\mathbf{P}_{t|t-1} = \mathbf{P}_{t|t-1}^* + \mathbf{P}$ and $\mathbf{B}_t = \mathbf{H} \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{C} \mathbf{R} \mathbf{C}^T$, we have:

$$\begin{aligned} \mathbf{B}_t &= \mathbf{A}_t \left(\mathbf{H} \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{C} \mathbf{R} \mathbf{C}^T \right) (\mathbf{A}_t)^T \\ &= \mathbf{A}_t \left(\mathbf{H} \mathbf{P}_{t|t-1}^* \mathbf{H}^T + \mathbf{B} \right) (\mathbf{A}_t)^T = \mathbf{B}_t^* \end{aligned} \quad (\text{B.4})$$

On the other hand $\mathbf{K}_t \mathbf{B}_t = \Phi \mathbf{P}_{t|t-1} \mathbf{H}^T (\mathbf{A}_t)^T + \mathbf{E} \mathbf{S} \mathbf{C}^T (\mathbf{A}_t)^T$ and, taking into account again that $\mathbf{P}_{t|t-1} = \mathbf{P}_{t|t-1}^* + \mathbf{P}$, we get:

$$\begin{aligned} \mathbf{K}_t \mathbf{B}_t &= \Phi \mathbf{P}_{t|t-1}^* \mathbf{H}^T (\mathbf{A}_t)^T + (\Phi \mathbf{P} \mathbf{H}^T + \mathbf{E} \mathbf{S} \mathbf{C}^T) (\mathbf{A}_t)^T \\ &= \Phi \mathbf{P}_{t|t-1}^* \mathbf{H}^T (\mathbf{A}_t)^T + \mathbf{K} \mathbf{B} (\mathbf{A}_t)^T = \mathbf{K}_t^* \mathbf{B}_t^* \end{aligned} \quad (\text{B.5})$$

Last, if we substitute (B.5) in (B.2) and take into account that $\mathbf{P}_{t|t-1} = \mathbf{P}_{t|t-1}^* + \mathbf{P}$ and $\mathbf{P} = \Phi \mathbf{P} \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T - \mathbf{K} \mathbf{B} (\mathbf{K})^T$, we obtain:

$$\begin{aligned}
\mathbf{P}_{t+1|t} &= \Phi \mathbf{P}_{t|t-1}^* \Phi^T + \Phi \mathbf{P} \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T - \mathbf{K}_t^* \mathbf{B}_t^* (\mathbf{K}_t^*)^T \\
&= \Phi \mathbf{P}_{t|t-1}^* \Phi^T + \mathbf{P} + \mathbf{K} \mathbf{B} \mathbf{K}^T - \mathbf{K}_t^* \mathbf{B}_t^* (\mathbf{K}_t^*)^T
\end{aligned} \tag{B.6}$$

and, as a consequence: $\mathbf{P}_{t+1|t} = \Phi \mathbf{P}_{t|t-1}^* \Phi^T + \mathbf{K} \mathbf{B} \mathbf{K}^T - \mathbf{K}_t^* \mathbf{B}_t^* (\mathbf{K}_t^*)^T + \mathbf{P} = \mathbf{P}_{t+1|t}^* + \mathbf{P}$ ■

Therefore, the presence of missing in-sample values changes the observer in both, the MEM and SEM forms, but the matrices \mathbf{B}_t and \mathbf{K}_t have the same structure and, therefore, the covariance matrices $\mathbf{P}_{t+1|t}$ and $\mathbf{P}_{t+1|t}^*$ maintain the same distance throughout all the sample.

The convergence of the SEM state covariances to zero cannot be proved in general. On the other hand, it is easy to check that it holds in particular cases, provided that the missing values are far enough from the end of the sample.

References

- [1] Anderson, B.D.O. and J.B. Moore, 1979, Optimal Filtering, Prentice-Hall, Englewood Cliffs (NJ).
- [2] Anderson, H.M. Low, C.M. and Snyder, R., 2006, Single source of error state space approach to the Beveridge Nelson decomposition, Economics Letters, 104-109.
- [3] Aoki, M. and A. Havenner, 1991, State Space Modeling of Multivariate Time Series, Econometric Reviews, 10, 1-59.
- [4] Bierman, G.J., 1977, Factorization Methods for Discrete Sequential Estimation, Prentice-Hall, Englewood Cliffs (NJ).
- [5] Box, G.E.P. Hillmer, S.C. and G.C. Tiao, 1978, Analysis and Modeling of Seasonal Time Series, in Seasonal Analysis of Time Series, A. Zellner (ed.) Bureau of the Census, Washington DC.
- [6] Casals, J. Sotoca, A. and M. Jerez, 1999, A Fast and Stable Method to Compute the Likelihood of Time Invariant State-Space Models, Economics Letters 65, 329-337.
- [7] Casals, J., M. Jerez and S. Sotoca, 2000, Exact Smoothing for Stationary and Nonstationary Time Series, International Journal of Forecasting 16, 1, 59-69.
- [8] Casals, J. and S. Sotoca, 2001, The Exact Likelihood for a State Space Model with Stochastic Inputs, Computers and Mathematics with Applications, 42, 199-209.
- [9] Casals, J. M. Jerez and S. Sotoca, 2002, An Exact Multivariate Model-based Structural Decomposition, Journal of the American Statistical Association, 97, 458, 553-564.
- [10] De Jong, P., 1989, Smoothing and Interpolation with the State Space Model, Journal of the American Statistical Association, 84, 408, 1085-1088.
- [11] De Jong, P. and S. Chu-Chun-Lin, 1994, Fast Likelihood Evaluation and Prediction for Nonstationary State Space Models, Biometrika, 81, 133-142.
- [12] De Jong, P. and J. Penzer, 2000, The ARIMA Model in State Space Form, London School of Economics, Department of Statistics, Research Report number 40, www.lse.ac.uk/statistics/documents/researchreport40.pdf
- [13] De Schutter, B., 2000, Minimal State-Space Realization in Linear System Theory: An Overview, Journal of Computational and Applied Mathematics, Special Issue on Numerical Analysis in the 20th Century - Vol. I: Approximation Theory, vol. 121, no. 1-2, 331-354.

- [14] De Souza, C.E. Gevers, M.R. and G.C. Goodwin, 1986, Riccati Equations in Optimal Filtering of Nonstabilizable Systems having Singular State Transition Matrices, IEEE Transactions on Automatic Control, AC-31, 9, 831-838.
- [15] Harvey, A.C., 1989, Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge (UK).
- [16] Harvey, A. and T. Trimbur, 2008, Trend Estimation and the Hodrick-Prescott Filter, Journal of the Japan Statistical Society, 38, 1, 41–49.
- [17] Hannan, E.J. and Deistler, M., 1988, The Statistical Theory of Linear Systems. John Wiley, New York.
- [18] Hodrick, R.J. and E.C. Prescott, 1997, Post-war U.S. Business Cycles: An Empirical Investigation, Journal of Money, Credit and Banking, 29, 1-16.
- [19] Kailath, T., 1980, Linear Systems. Prentice–Hall, Englewood Cliffs.
- [20] Morley, J.A., 2010, The Two Interpretations of the Beveridge Nelson Decomposition, Macroeconomic Dynamics, 1-21.
- [21] Ord, J.K. Koehler, A.B. and R.D. Snyder, 1997, Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models. Journal of the American Statistical Association, 92, 440, 1621-1629.
- [22] Park, P.G. and T. Kailath, 1997, Convergence of the DRE Solution to the ARE Strong Solution, IEEE Transactions on Automatic Control, 42, 573-578.
- [23] Proietti T., 2006, Trend–Cycle Decompositions with Correlated Components. Econometric Reviews, 25, 61-84.
- [24] Terceiro, J., 1990, Estimation of Dynamic Econometric Models with Errors in Variables, Springer-Verlag, Berlin.

Table 4.1: Structural time series model. As usual, all the errors are assumed to be independent. The second column shows the ARIMA models for each component.

Component	Structural representation	ARIMA representation
Time series	$z_t = \mu_t + \gamma_t + e_t$, $e_t \sim \text{iid}(0,1)$	$(1-B)(1-B^4)z_t$ $= (1-.933B+.091B^2-.047B^3-.585B^4+.548B^5)a_t$ $\simeq (1-.933B)(1-.585B^4)a_t$; $a_t \sim \text{iid}(0,1.824)$
Trend	$\mu_{t+1} = \mu_t + \beta_t$ $\beta_{t+1} = \beta_t + \zeta_t$, $\zeta_t \sim \text{iid}(0,1/1600)$	$(1-B)^2 \mu_{t+1} = \zeta_t$
Seasonal	$\gamma_{t+1} = \gamma_t + \gamma_{t-1} + \gamma_{t-2} + \omega_t$ $\omega_t \sim \text{iid}(0, .1)$	$(1+B+B^2+B^3)\gamma_{t+1} = \omega_t$

Table 4.2. Structure of the SEM. The variance of the error term is $\hat{\sigma}_a^2 = 1.824$. The transition matrix is structured in block-diagonal form to simplify its interpretation. Accordingly, the trend is given by the first state and the seasonal component is a combination of the third, fourth and fifth states with the corresponding coefficients in the observation equation.

Outputs	Inputs					
	x_{1t}^*	x_{2t}^*	x_{3t}^*	x_{4t}^*	x_{5t}^*	a_t
	Φ					K
x_{1t+1}^*	1	1	0	0	0	.188
x_{2t+1}^*	0	1	0	0	0	.019
x_{3t+1}^*	0	0	.489	1.461	0	-.070
x_{4t+1}^*	0	0	-.848	-.489	0	-.116
x_{5t+1}^*	0	0	0	0	-1	-.203
z_t	H					—
	1	0	.619	-.342	-.577	1

Figure 3.1 Convergence of the trace of the smoothing covariance, $\text{tr}(\mathbf{P}_{t|N})$, for different model structures. The calculations have been done using 50 simulated samples of the process indicated in each case. The letter B denotes the backshift operator, such that for any sequence y_t : $B^k y_t = y_{t-k}$, $k = 0, \pm 1, \pm 2, \dots$

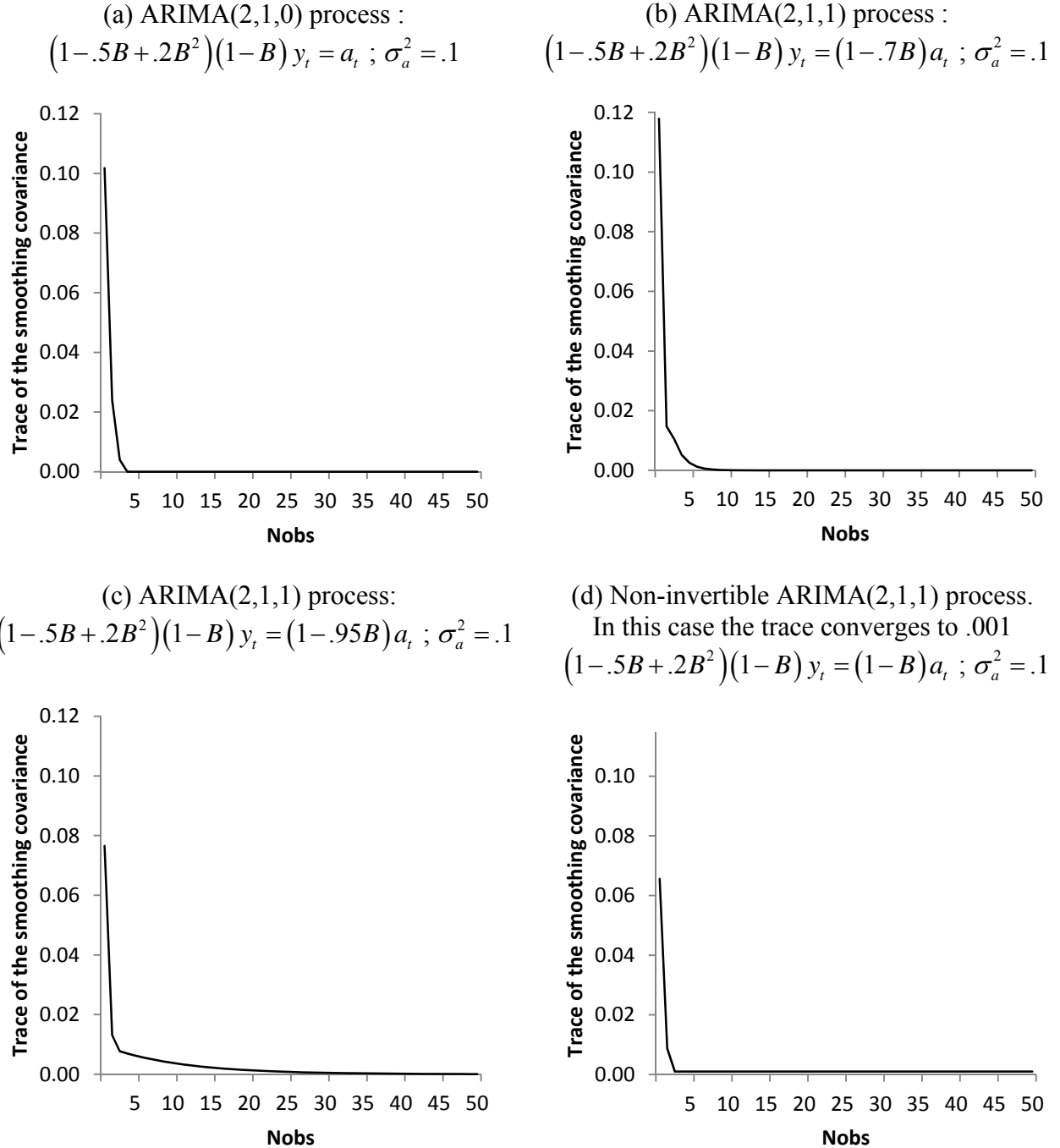


Figure 4.1.a. Traces of the smoother covariances corresponding to the MEM (thin line) and SEM (thick line).

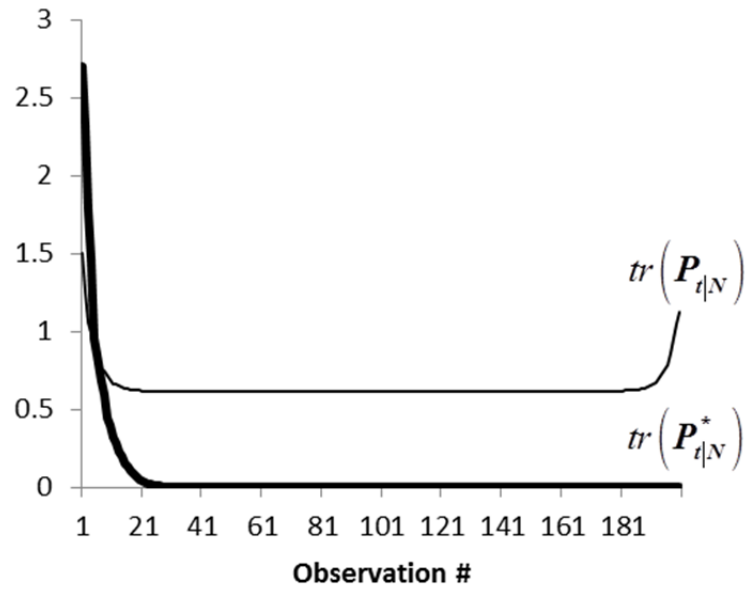


Figure 4.1.b. Traces of the smoothed state covariances corresponding to the MEM (thin line) and SEM (thick line), computed from a sample with missing values in observations 100 to 104.

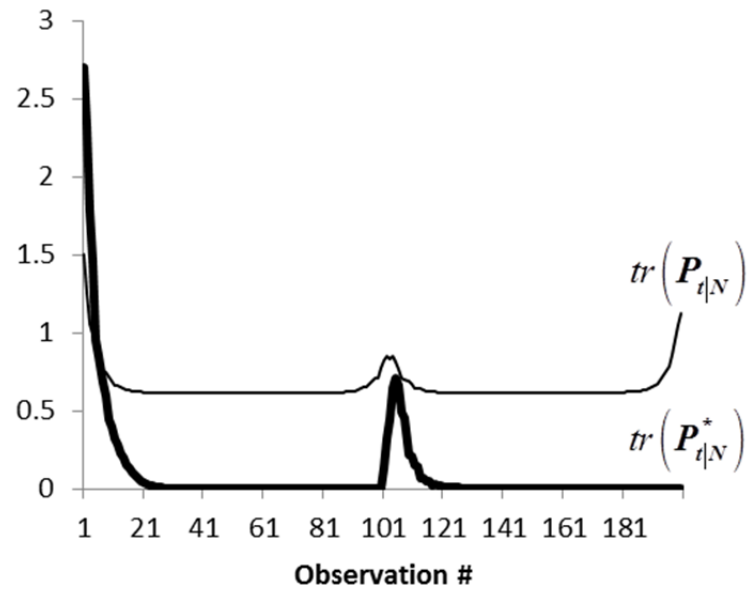


Figure 4.2.a. Sample autocorrelation function of the four-quarters rolling sum of the MEM seasonal component. The dashed lines are the approximate 95% significance bounds for the coefficients. Seasonal lags are emphasized with hollow bars.

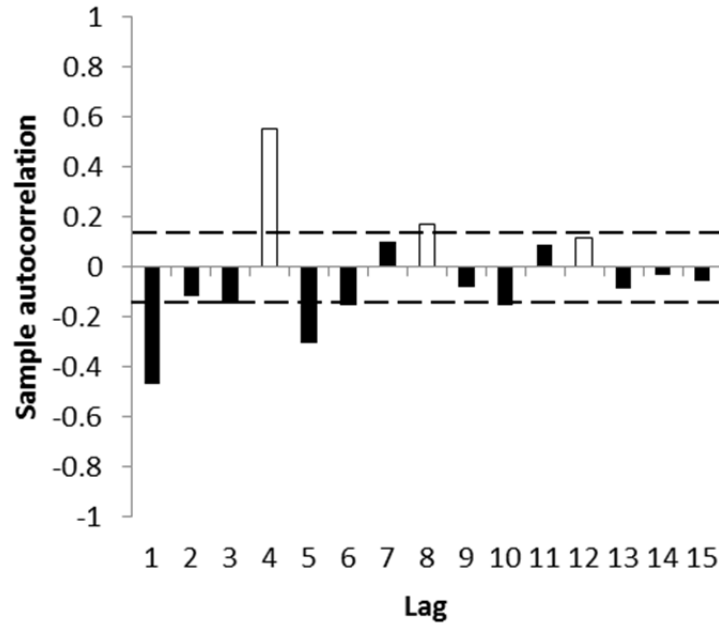


Figure 4.2.b. Sample autocorrelation function of the four-quarters moving sum of the SEM seasonal component.

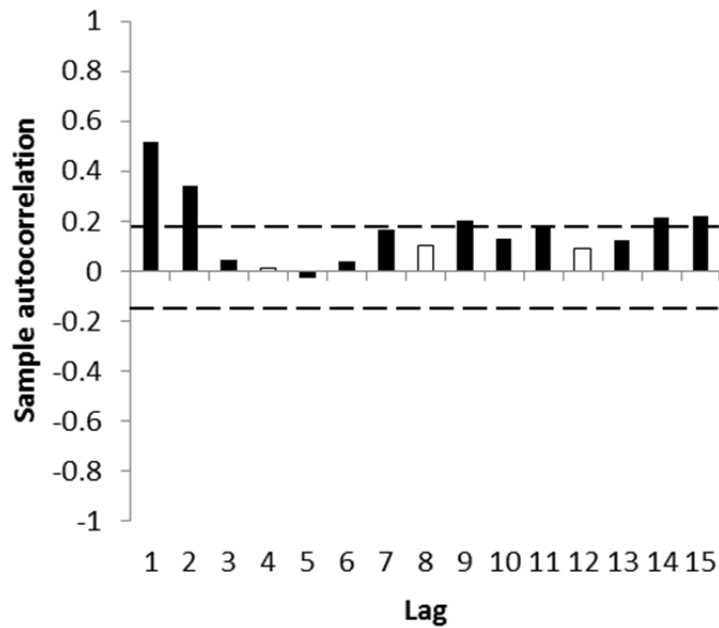


Figure 4.3 Change in the MEM trend (thick line) vs. change in the SEM (thin line).

